Mobile Robot Localization via Machine Learning

Alexander Kuleshov[1], Alexander Bernstein[1,2] and Evgeny Burnaev[1,2]

[1] Skolkovo Institute of Science and Technology, Moscow, Russia
A.Kuleshov@skoltech.ru, A.Bernstein@skoltech.ru, E.Burnaev@skoltech.ru
[2] Kharkevich Institute for Information Transmission Problems RAS, Moscow, Russia

Abstract. We consider an appearance-based robot self-localization problem in the machine learning framework. Using recent manifold learning techniques, we propose a new geometrically motivated solution. The solution includes estimation of the robot localization mapping from the appearance manifold to the robot localization space, as well as estimation of the inverse mapping for image modeling. The latter allows solving the robot localization problem as a Kalman filtering problem.

Keywords: Machine learning; Robotics; Mobile robot self-localization; Appearance-based learning; Manifold learning; Regression on manifolds

1 Introduction

Machine learning is an essential and ubiquitous framework for solving a wide range of tasks in different application areas. One of these tasks is the problem of estimating position (localization) of a mobile robot moving in an uncertain environment, which is necessary for understanding the environment to make navigational decisions.

1.1 Robot Localization Problem

Localization is a fundamental component of mobile robotics to perform certain tasks in a given workspace. If a robot does not know where it is, it cannot navigate effectively and achieve goals. In order for a mobile robot to travel from one location to another it has to know its position and orientation (pose) at any given moment of time. Estimation of a mobile robot pose (localization) with respect to known locations in an environment is called the localization problem.

This task is especially important for navigation of autonomous robots when the localization problem is solved with the use of sensor systems installed on the robot. In this case, the task is called the self-localization problem.

The most common and basic method for performing localization is through dead-reckoning using data received from odometer sensors. This technique integrates a history of sensor readings and executed actions (e.g., velocity history) of the robot over time to determine changes in positions from its starting location ([1, 2]). Unfortunately, pure dead-reckoning methods are prone to errors growing over time, so we need to periodically adjust the robot position with the help of other sensor systems, e.g. a visual sensing system (cameras). Usually in order to probabilistically update the robot position [4] an extended Kalman filter [3] is used.

Passive vision-based localization methods, including methods based only on visual data, have received extensive attention over the past few decades from the robotic and machine vision communities. Surveys [5, 6] describe approaches to visual navigation of autonomous mobile robots, whose performance significantly depends on accurate robot localization in an environment. Thus vision-based localization solutions, developed for autonomous robots, are reviewed in these surveys.

We consider the self-localization problem of a mobile robot as a specific machine vision problem in which the robot position is estimated from images captured by its visual system. Omnidirectional imaging system, consisting of a vertically oriented standard color camera and a hyperbolic mirror, mounted in front of the lens [7–10] or camera with steerable orientation [11], is a typical example of such visual system.

Continuous set of images, which can be captured by a camera under all possible image formation parameters, is called the Appearance space. Parameters, defining the appearance space, involve relative position and orientation of the robot moving in a certain workspace, as well the camera intrinsic parameters and its illumination function (including color, intensity and angle). In this paper, we consider only the most typical case when these parameters contain robot localization (pose) consisting of robot position and orientation only.

Assuming that captured images allow distinguishing and recognizing poses from which they have been taken, the solution to the considered appearance-based (passive vision-based) robot localization problem is a mapping from the Appearance space to the Localization space consisting of all possible robot localizations.

1.2 Robot Localization: Machine Learning Framework

We consider the appearance-based localization problem in the machine learning framework: the appearance-based model describes the relation between observed images and their locations; the model is constructed using training data, which consists of captured images taken in known positions; the model allows estimating an unknown robot position from a newly acquired image. Such appearance-based learning framework has become very popular in the field of robot learning [12].

Because images are represented as very high-dimensional vectors, various appearance-based models (aka maps), describing the underlying low-dimensional structure in the appearance space, are usually constructed from training positions-images data using supervised learning techniques [13–16]. Appearance-based models provide an internal representation of the appearance space consisting of certain visual features, extracted from images. A number of methods for building such models using sonar, odometer and optical sensors have been proposed in the last few years. Given such model, the robot localization task can be considered as a prediction problem: to predict position of a robot from a new sensor observation as accurately as possibly.

1.3 Related Works

A geometrical approach to Machine Vision is to extract and estimate relative positions of detected environmental features or landmarks (e.g., doors, corners, columns, etc. in office workspace) [17, 18] and then track over time their three-dimensional positions in order to estimate robot ego-motion [19, 20].

In some cases, visual landmarks (as well as objects) are recognized by projecting rectangular sub-regions from newly acquired images into a space of descriptors and associating them with nearby ones. Various methods for recognizing landmark objects and scenes in images by a robot visual system are described in [21–27].

Appearance-based methods consider input images holistically, in relation to other images. Usually in case of appearance-based methods, as well as in geometrical approaches [23–27], natural visual features are computed by projecting images or range data onto low-dimensional subspaces [11, 28–31]. For that Principal Component Analysis (PCA) [32] is often used. A review of PCA and other related subspace methods for the appearance-based localization problem is provided in [10]. There are a number of methods to determine significant variables in PCA. According to [33], the merit of these methods highly depends on the extent to which variables are correlated.

A mapping from visual features to the robot pose is highly nonlinear and sensitive to the type of selected features. Since PCA-features might not be able to maximize localization performance, other criteria are used to select linear subspaces [7–9]. For example, the subspace can be selected via Canonical Correlation Analysis (CCA) [34] to maximize correlations between poses of the robot and captured images [35, 36]. In contrast to PCA, CCA performs a sort of a regression task in a sense that it provides a one-to-one mapping between an image and a camera pose [9]. A closely related problem is investigated in [37, 29], in which ego-motion technique is used for estimating the robot pose based on Scale-Invariant Features Transform (SIFT features) [38] and Gaussian process regression [63–65]. In [39] linear projections of supervised high-dimensional robot observations are constructed to minimize conditional entropy of robot positions given the projected observations. Iterative technique for the minimization problem is proposed, which starts from the CCA-solution.

In [10] a probabilistic localization algorithm is proposed, which directly maps high-dimensional appearance images to robot positions via a nonlinear dimensionality reduction. Starting from Locally Linear Embedding manifold learning algorithm [40], Locally Linear Projection (LLP) method with subsequent usage of Bayesian filtering technique is proposed for robot pose estimation.

Regression dependency between PCA-based features of images and robot coordinates is constructed in [41] on the basis of the proposed General regression neural network, trained on panoramic snapshots.

Simultaneous Localization And Mapping (SLAM) algorithm is presented in [43], allowing to get a real-time localization. The algorithm combines dimension reduction methods (kernel principal component analysis [42]), manifold regularization techniques and parameter selection.

In [44] a multivariate Gaussian process (GP) with unknown hyper-parameters, estimated using maximum likelihood technique, models sought-for visual features of the data from an omnidirectional camera. Then GP regression is used for estimating the robot localization.

Papers [45, 46] present appearance-based localization for an omnidirectional camera that builds on a combination of the group Least Absolute Shrinkage and Selection Operator (LASSO) [47] and the Extended Kalman filter (EKF) [3, 48].

Speeded-Up Robust Features (SURF points) [49], being scale- and rotation-invariants utilizing Haar wavelet responses to produce a 64-dimensional descriptor vector for points of interest in an image, are computed for each image via the group LASSO regression [50]. The EKF uses an output of the LASSO regression-based initial localization as observations for a final localization.

1.4 Paper contribution

The paper proposes new geometrically motivated manifold learning approach [51] to the solution of the appearance-based robot localization problem using specific Grassmann & Stiefel Eigenmaps algorithm [52] for nonlinear dimensionality reduction and regression on manifolds [53].

The paper is organized as follows. Section 2 contains rigorous statement of the appearance-based robot localization problem. Proposed approach is described in Section 3. Section 4 provides details of this solution. Section 5 describes how to use the solution in Kalman filtering procedures for robot localization.

2 Robot Localization: Rigorous Problem Statement

Let a mobile robot, equipped with a visual system (for example, an omnidirectional imaging system), moves on a 2D-workspace $\mathbf{W} \subset \mathbb{R}^2$. Its localization $\boldsymbol{\theta} = (\theta_{RC}, \theta_{RO}) \in \mathbb{R}^3$ is a three-dimensional vector consisting of **R**obot **P**osition $\theta_{RP} \in \mathbf{W} \subset \mathbb{R}^2$ (robot coordinates in the workspace $\mathbf{W}$) and **R**obot **O**rientation (an angle) $\theta_{RO} \in \mathbb{R}^1$ relative to the coordinate system in the workspace. Note that in case of a camera with steerable orientation, **C**amera **O**rientation (an angle) $\theta_{CO} \in \mathbb{R}^1$ relative to the robot should be included in the robot localization parameter, but for simplicity we consider only three-dimensional robot localization parameter $\boldsymbol{\theta}$. Let us denote by $\boldsymbol{\Theta} \subset \mathbb{R}^3$ a subset consisting of all possible robot localization parameters and called Localization space.

Let an image, captured by the robot imaging system, consists of p pixels. Therefore we can represent the image by a p-dimensional image-vector X. We denote by $X = \varphi(\boldsymbol{\theta}) \in \mathbb{R}^p$ an image, captured by the robot with the localization parameter $\boldsymbol{\theta}$, which is described by the Image modeling function φ with the domain of definition $\boldsymbol{\Theta}$. Let

$$\mathbf{X} = \varphi(\boldsymbol{\Theta}) = \{X = \varphi(\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^3\} \subset \mathbb{R}^p, \tag{1}$$

be an Appearance space consisting of all possible images, which can be captured by the mobile robot and parameterized by the robot localization parameter $\boldsymbol{\theta} \in \mathbb{R}^3$.

We assume that images, captured by the robot in different localizations, are different; hence, the Image modeling function $\varphi : \boldsymbol{\Theta} \rightarrow \mathbf{X}$ is a one-to-one mapping from the Localization space to the Appearance space. Thus the Appearance space $\mathbf{X}$ is a manifold (Appearance manifold, AM) without self-intersections and with intrinsic dimension $q = 3$. This manifold is parameterized by the chart φ and is embedded in the ambient p-dimensional Euclidean space. Therefore, there exists an inverse mapping

$$\psi = \varphi^{-1} : X \in \mathbf{X} \rightarrow \boldsymbol{\Theta} = \psi(X) \in \boldsymbol{\Theta}, \tag{2}$$

called Localization function from the AM $\mathbf{X}$ to the Localization space $\boldsymbol{\Theta}$.

The functions φ and ψ, as well the AM $\mathbf{X}$, are unknown, and the Robot localization problem consists in constructing the robot localization $\boldsymbol{\theta} = \psi(X)$ from the image $X = \varphi(\boldsymbol{\theta}) \in \mathbf{X}$. We consider this problem in the machine learning framework [12]. We have a training set

$$\mathbf{S}_n = \{(X_i, \boldsymbol{\theta}_i), i = 1, 2, \ldots, n\}, \tag{3}$$

consisting of images $\{X_i = \varphi(\boldsymbol{\theta}_i)\}$, captured by the robot imaging system in known conditions $\{\theta_i \in \boldsymbol{\Theta}\}$ when robot moves in the workspace $\mathbf{W}$ randomly or on a regular grid. For example, the mobile robot, described in [7, 8], captures omnidirectional images every 25 centimeters along robot random paths; reference positions are located on a regular grid with cells of size either 25cm in a 2.7m×5.4m workspace [9] or 1m in a 20m×20m workspace [10], respectively.

We consider two interconnected statements related to the robot localization problem and based on two representations of the training dataset $\mathbf{S}_n$ (3):

– Estimating the Localization function problem: to recover an unknown Localization function $\boldsymbol{\theta} = \psi(X)$ at an arbitrary out-of-sample point $X \in \mathbf{X}$, from its known values $\{\boldsymbol{\theta}_i = \psi(X_i)\}$ at known points $\{X_i\}$,
– Estimating the Image modeling function problem: to predict a captured image $X = \varphi(\boldsymbol{\theta})$ at an arbitrary out-of-sample point $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ from the known images $\{X_i = \varphi(\boldsymbol{\theta}_i)\}$ captured at known points $\boldsymbol{\theta}_i$.

The solution to the inverse problem can be used in the incremental statement of the robot localization problem considered below.

Both considered problems are regression problems with high-dimensional manifold valued inputs and high-dimensional manifold valued outputs, respectively.

Most of known appearance-based learning methods are related only to the first problem "Estimating the Localization function". In this regression problem the dimensionality p of input vectors $X \in \mathbf{X}$ is large, for example, $p =$

16384, 10240, and 3925 in case of panoramic images considered in [7–10], respectively; $p = 6912$ and $p = 4096$ in two examples considered in [11]. Thus standard regression methods perform poorly due the statistical and computational "curse of dimensionality" phenomenon: collinearity or "near-collinearity" of high-dimensional inputs cause difficulties when constructing regression; regression error cannot achieve a convergence rate faster than $n^{-s/(2s+p)}$ when estimating at least s times differentiable function $f(X)$ [54, 55].

In order to avoid the curse of dimensionality phenomena, we use the fact that the Appearance space $\mathbf{X}$ (1) is a nonlinear manifold with intrinsic dimensionality $q = 3$, and q-dimensional visual features are constructed and used in the proposed method. Note that most of conventional methods use various dimensionality reduction techniques (usually, PCA) for constructing low-dimensional visual features by projecting images onto constructed low-dimensional subspaces, and the robot localization function is estimated from these features. However constructed low-dimensional subspaces may have a larger dimension than the real intrinsic dimension $q = 3$, i.e. they are not optimal.

3 Robot Localization: Proposed Approach

3.1 Regression Manifold

Consider an unknown smooth manifold called Regression manifold (RM)

$$\mathbf{M} = \{Z = F(\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^3\} \subset \mathbb{R}^{p+3} \tag{4}$$

with the intrinsic dimension $q = 3$, which is embedded in an ambient $(p+3)$-dimensional Euclidean space and parameterized by an unknown chart

$$F : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^3 \to Z = F(\boldsymbol{\theta}) = \begin{pmatrix} \varphi(\boldsymbol{\theta}) \\ \boldsymbol{\theta} \end{pmatrix} \in \mathbf{M}, \tag{5}$$

defined on the Localization space $\boldsymbol{\Theta}$. Let

$$J_F(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}) = \begin{pmatrix} J_\varphi(\boldsymbol{\theta}) \\ \mathbf{I}_3 \end{pmatrix}, \tag{6}$$

be $(p+3) \times q$ Jacobian matrix of the mapping F (5) which is split into $p \times 3$ Jacobian matrix $J_\varphi(\boldsymbol{\theta})$ of the Image modeling function φ (1) and $\mathbf{I}_3$ being a 3×3 unit matrix. The Jacobian $J_F(\boldsymbol{\theta})$ (6) determines a three-dimensional linear space $L(Z) = \mathrm{Span}(J_F(\boldsymbol{\theta}))$ in $\mathbb{R}^{p+3}$ which is a tangent space to the RM $\mathbf{M}$ at the point $Z = F(\boldsymbol{\theta}) \in \mathbf{M}$; hereinafter, $\mathrm{Span}(H)$ is a linear space spanned by columns of an arbitrary matrix H.

The set $\mathrm{TB}(\mathbf{M}) = \{(Z, L(Z)) : Z \in \mathbf{M}\}$ composed of points Z of the RM $\mathbf{M}$ equipped by tangent spaces $L_F(X)$ at these points is known in the Manifold theory [56, 57] as the Tangent Bundle of the RM $\mathbf{M}$.

3.2 Tangent Bundle Manifold Learning problem for Regression Manifold

The dataset $\mathbf{S}_n$ (3), written in the form

$$\begin{aligned} \mathbf{S}_n &= \left\{ Z_i = \begin{pmatrix} X_i = \varphi(\boldsymbol{\theta}_i) \\ \boldsymbol{\theta}_i \end{pmatrix}, i = 1, 2, \ldots, n \right\} \\ &= \left\{ Z_i = \begin{pmatrix} X_i \\ \boldsymbol{\theta}_i = \psi(X_i) \end{pmatrix}, i = 1, 2, \ldots, n \right\}, \end{aligned} \tag{7}$$

can be considered as a sample from the unknown RM $\mathbf{M}$ (4).

Let us consider certain dimensionality reduction problem called Tangent bundle manifold learning problem [58, 59] for the RM $\mathbf{M}$: estimate the Tangent Bundle TB($\mathbf{M}$) given the sample $\mathbf{S}_n$ (7) from the unknown RM $\mathbf{M}$.

3.3 Grassmann & Stiefel Eigenmaps solution

Using Grassmann & Stiefel Eigenmaps (GSE) method [58, 59] and the sample $\mathbf{S}_n$ we construct the solution to the Tangent bundle manifold learning problem, resulting in the following quantities:

- sample-based area $\mathbf{M}^* \subset \mathbb{R}^{p+3}$ which is close to the unknown RM $\mathbf{M}$,
- embedding mapping $h_{GSE}(Z)$ from the area $\mathbf{M}^*$ to the 3-dimensional Feature space (FS) $\mathbf{Y}_{GSE} = h_{GSE}(\mathbf{M}^*) \subset \mathbb{R}^3$,
- recovery mapping $g_{GSE}(y)$ from the FS $\mathbf{Y}_{GSE}$ to $\mathbb{R}^{p+3}$,
- $(p+3) \times q$ matrix $G_{GSE,g}(y)$ defined on the FS $\mathbf{Y}_{GSE}$,

which together provides both

- proximity

  $$Z_{GSE}(Z) \equiv g_{GSE}(h_{GSE}(Z)) \approx Z \quad \text{for all points} \quad Z \in \mathbf{M}^*, \tag{8}$$

  between initial and recovered points Z and $Z_{GSE}(Z)$. Thanks to (8) we get small Hausdorff distance $d_H(\mathbf{M}, \mathbf{M}_{GSE})$ between the RM $\mathbf{M}$ and the three-dimensional recovered regression manifold (RRM)

  $$\mathbf{M}_{GSE} = \{g_{GSE,g}(y) \in \mathbb{R}^{p+3} : y \in \mathbf{Y}_{GSE} \subset \mathbb{R}^3\}, \tag{9}$$

  embedded in the ambient $(p+3)$-dimensional Euclidean space;
- proximity

  $$G_{GSE,g}(y) \approx J_{GSE,g}(y) \quad \text{for all points} \quad y \in \mathbf{Y}_{GSE}, \tag{10}$$

  in which $J_{GSE,g}(y)$ is a Jacobian matrix of the mapping $g_{GSE}(y)$. Thanks to (10) we get proximity between the tangent space $L(Z)$ to the RM $\mathbf{M}$ at the point Z and the tangent space $L_{GSE}(Z) = \mathrm{Span}(G_{GSE,g}(h_{GSE}(Z)))$ to the RRM $\mathbf{M}_{GSE}$ (9) at the nearby recovered point $Z_{GSE}(Z)$. The proximity between these tangent spaces, considered as elements of the Grassmann manifold, is defined using chosen metric on the Grassmann manifold.

Therefore, the tangent bundle $\mathrm{TB}(\mathbf{M}_{GSE}) = \{(Z_{GSE}(Z), L_{GSE}(Z)) : Z \in \mathbf{M}\}$ of the RRM $\mathbf{M}_{GSE}$ accurately approximates the tangent bundle $\mathrm{TB}(\mathbf{M})$.

Note also that the original GSE algorithm [58, 59] has computational complexity $O(n^3)$ for a sample size n; the incremental version of the GSE [60] has significantly smaller running time $O(n^{(q+4)/(q+2)})$.

3.4 Robot Localization: GSE-based approach

A splitting of an arbitrary vector $Z = \begin{pmatrix} Z_u \\ Z_v \end{pmatrix} \in \mathbb{R}^{p+3}$ into two vectors $Z_u \in \mathbb{R}^p$ and $Z_v \in \mathbb{R}^3$ implies the corresponding partitions

$$g_{GSE}(y) = \begin{pmatrix} g_{GSE,u}(y) \\ g_{GSE,v}(y) \end{pmatrix}, G_{GSE,g}(y) = \begin{pmatrix} g_{GSE,g,u}(y) \\ g_{GSE,g,v}(y) \end{pmatrix} \tag{11}$$

of the mapping $g_{GSE}(y)$ and the matrix $G_{GSE,g}(y)$.

Using the representation $Z = F(\boldsymbol{\theta})$ (5), the embedding mapping $y = h_{GSE}(Z)$, defined on the RM $\mathbf{M}$, can be written as a function

$$y = R_{GSE}(\boldsymbol{\theta}) = h_{GSE}(F(\boldsymbol{\theta})), \tag{12}$$

defined on the Localization space $\boldsymbol{\Theta}$.

Using the Localization function ψ (2), the RM $\mathbf{M}$ and the embedding mapping $y = h_{GSE}(Z)$ can be written as $\mathbf{M} = \{Z = f(X),\ X \in \mathbf{X}\} \subset \mathbb{R}^{p+3}$ and

$$y = r_{GSE}(X) = h_{GSE}(f(X)) \tag{13}$$

respectively, where the functions

$$f(X) = \begin{pmatrix} X \\ \psi(X) \end{pmatrix} \tag{14}$$

and $r_{GSE}(X)$ (13) are defined on the Appearance space $\mathbf{X}$.

In the paper [53], the $3 \times p$ and 3×3 Jacobian matrices $J_{GSE,r}(X)$ (the covariant differentiation is used here) and $J_{GSE,R}(\boldsymbol{\theta})$ of the mappings $r_{GSE}(X)$ and $R_{GSE}(\boldsymbol{\theta})$ are estimated by the matrices

$$G_{GSE,r}(X) = G^{-}_{GSE,g,u}(r_{GSE}(X)) \times \pi_{GSE,\mathbf{X}}(X), \tag{15}$$

$$G_{GSE,R}(\boldsymbol{\theta}) = G^{-1}_{GSE,g,v}(R_{GSE}(\boldsymbol{\theta})), \tag{16}$$

respectively. Here $H^{-} = (H^{\mathrm{T}} \times H)^{-1} \times H^{\mathrm{T}}$ denotes a pseudo-inverse Moore-Penrose matrix [61] of an arbitrary matrix H and $\pi_{GSE,\mathbf{X}}(X)$ is a certain estimator [62] of a $p \times p$ projection matrix onto the tangent space of the AM $\mathbf{X}$ (1) at the point $X \in \mathbf{X}$.

Using representations (12) and (13), the proximity (8), implies approximate equalities

$$\psi_{GSE}(X) \equiv g_{GSE,v}(r_{GSE}(X)) \approx \psi(X), \tag{17}$$

$$\varphi_{GSE}(\boldsymbol{\theta}) \equiv g_{GSE,u}(R_{GSE}(\boldsymbol{\theta})) \approx \varphi(\boldsymbol{\theta}). \tag{18}$$

Although the GSE-based functions $\psi_{GSE}(X)$ (17) and $\varphi_{GSE}(\boldsymbol{\theta})$ (18) accurately approximate the sought-for functions $\psi(X)$ and $\varphi(\boldsymbol{\theta})$, respectively, they cannot be considered as the solution to the Robot Localization problem because the mappings $g_{GSE,u}(y)$ and $g_{GSE,v}(y)$ (11) depend on the argument

$$y = r_{GSE}(X) = R_{GSE}(\boldsymbol{\theta}), \tag{19}$$

whose values are known only at sample points:

$$y_i = r_{GSE}(X_i) = R_{GSE}(\boldsymbol{\theta}_i) = h_{GSE}(Z_i), i = 1, 2, \ldots, n. \tag{20}$$

Based on known values (20) of the functions $r_{GSE}(X)$ and $R_{GSE}(\boldsymbol{\theta})$ (19) at sample points, as well on the known values of their Jacobian matrices (15), (16) at these points, the estimators $r^*(X)$ and $R^*(\boldsymbol{\theta})$ of these functions at arbitrary points $X \in \mathbf{X}$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, respectively, can be constructed using the Jacobian Regression method, proposed in [53]. Substitution of estimators $r^*(X)$ and $R^*(\boldsymbol{\theta})$ in formulas (17) and (18) instead of $r_{GSE}(X)$ and $R_{GSE}(\boldsymbol{\theta})$, provides the final estimators $\psi^*(\mathbf{X})$ and $\varphi^*(\boldsymbol{\theta})$ for the Localization function $\psi(X)$ and the Image modeling function $\varphi(\boldsymbol{\theta})$.

4 Robot Localization: Solution

4.1 Estimating the Localization function

GSE solution applied to the RM $\mathbf{M}$ includes construction of the kernels $K_X(X, X')$ and $K_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ on the AM $\mathbf{X}$ and the Localization space $\boldsymbol{\Theta}$, respectively. These kernels reflect not only geometrical closeness between points $Z = F(\boldsymbol{\theta}) = f(X)$ and $Z' = F(\boldsymbol{\theta}') = f(X')$ but also closeness between the tangent spaces $L(Z)$ and $L(Z')$ to the RM $\mathbf{M}$.

Estimating the GSE-based embedding mapping $r_{GSE}(X)$. For close points $X, X' \in \mathbf{X}$, write the Taylor series expansion

$$r_{GSE}(X) - r_{GSE}(X') \approx J_{GSE,r}(X') \times (X - X')$$

for the embedding function $y = r_{GSE}(X)$.

Taking as points X' the sample images $\{X_i\}$, using known values y_i (20) of $r_{GSE}(X_i)$, and replacing the Jacobians $J_{GSE,r}(X_i)$ by estimators $\{G_{GSE,r}(X_i)\}$ (15) known at sample points, construct the estimator $y^* = r^*(X)$ for the embedding mapping $y = r_{GSE}(X)$ as a minimizer of the cost function

$$\sum_{i=1}^{n} K_X(X, X_i) \times |y - y_i - G_{GSE,r}(X_i) \times (X - X_i)|^2$$

over y. The solution $r^*(X)$ to this problem is computed in an explicit form

$$r^*(X) = \frac{1}{K_X(X)} \times \sum_{i=1}^{n} K_X(X, X_i) \times \{y_i + G_{GSE,r}(X_i) \times (X - X_i)\}, \tag{21}$$

where $K_X(X) = \sum_{i=1}^{n} K_X(X, X_i)$.

Final estimator for the Localization function at arbitrary point $X \in \mathbf{X}$ is given by the formula

$$\psi^*(X) \equiv g_{GSE,v}(r^*(X)). \tag{22}$$

4.2 Estimating the Image modeling function

Estimating the GSE-based embedding mapping $R_{GSE}(\boldsymbol{\theta})$. For close points $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$, write the Taylor series expansion

$$R_{GSE}(\boldsymbol{\theta}) - R_{GSE}(\boldsymbol{\theta}') \approx J_{GSE,R}(\boldsymbol{\theta}') \times (\boldsymbol{\theta} - \boldsymbol{\theta}')$$

for the embedding function $y = R_{GSE}(\boldsymbol{\theta})$.

Taking as points $\boldsymbol{\theta}'$ the sample localizations $\{\boldsymbol{\theta}_i\}$, using known values $\{y_i\}$ (20) of $\{R_{GSE}(\boldsymbol{\theta}_i)\}$, and replacing the Jacobians $\{J_{GSE,R}(\boldsymbol{\theta}_i)\}$ by their estimators $\{G_{GSE,R}(\boldsymbol{\theta}_i)\}$ (16) known at sample points, choose the estimator $y^* = R^*(\boldsymbol{\theta})$ for the embedding mapping $y = R_{GSE}(\boldsymbol{\theta})$ as a quantity that minimizes the cost function

$$\sum_{i=1}^{n} K_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\theta}_i) \times |y - y_i - G_{GSE,R}(\boldsymbol{\theta}_i) \times (\boldsymbol{\theta} - \boldsymbol{\theta}_i)|^2$$

over y. The solution $R^*(\boldsymbol{\theta})$ to this problem is computed in an explicit form

$$R^*(\boldsymbol{\theta}) = \frac{1}{K_{\boldsymbol{\theta}}(\boldsymbol{\theta})} \times \sum_{i=1}^{n} K_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\theta}_i) \times \{y_i + G_{GSE,R}(\boldsymbol{\theta}_i) \times (\boldsymbol{\theta} - \boldsymbol{\theta}_i)\}, \tag{23}$$

where $K_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} K_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\theta}_i)$.

Final estimator for the Image modeling function at arbitrary point $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is given by the formula

$$\varphi^*(\boldsymbol{\theta}) \equiv g_{GSE,u}(R^*(\boldsymbol{\theta})). \tag{24}$$

5 Usage of Predicted Images for Robot Localization

The constructed estimator $\varphi^*(\boldsymbol{\theta})$ (24), which predicts an image $X = \varphi(\boldsymbol{\theta})$ captured at the localization $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, can be used in Kalman filtering procedures (3) for robot localization.

Robot navigation consists in choosing control $u(t)$ at given time moments $t = 0, 1, 2, \ldots$. Let $\boldsymbol{\theta}(t)$ be a current robot position at time t, then, under some control $u(t)$, the robot must move in the expected position

$$\boldsymbol{\theta}(t+1) = F(\boldsymbol{\theta}(t), u(t)) \equiv F_t(\boldsymbol{\theta}(t)), \tag{25}$$

where $F(\boldsymbol{\theta}, u)$ is a known function defined by a solution of a navigation motion control problem.

In practice, the estimated position $\boldsymbol{\theta}_t(t)$ of the robot at time t is only known, which differs from the exact position $\boldsymbol{\theta}(t)$; the exact position $\boldsymbol{\theta}(t+1)$ at time $(t+1)$ also differs from the expected position $\boldsymbol{\theta}_t(t+1) = F_t(\boldsymbol{\theta}_t(t))$.

Let a robot visual sensing system provides a captured image $X(t+1) = \varphi(\boldsymbol{\theta}(t+1))$ at the moment $(t+1)$. We want to solve a filtering problem to improve the predicted localization $\boldsymbol{\theta}_t(t+1)$ from the captured image $X(t+1)$.

The constructed estimator $\varphi^*(\boldsymbol{\theta})$ (24) allows predicting $X^*(t+1) = \varphi^*(\boldsymbol{\theta}_t(t+1))$ for the captured image $X(t+1)$, and a standard Kalman filter [3] constructs the improved localization $\boldsymbol{\theta}_{t+1}(t+1)$ as

$$\boldsymbol{\theta}_{t+1}(t+1) = \boldsymbol{\theta}_t(t+1) + B(t+1) \times (X(t+1) - X^*(t+1)). \qquad (26)$$

Here $B(t+1)$ is a Kalman gain.

Using the estimator $\psi^*(X)$ (22) for the Localization function $\psi(X)$ (2), we can use a quantity $\psi^*(X(t+1))$, representing visual features, as an estimator of the robot pose in which the image $X(t+1)$ has been taken, and construct the estimator $\boldsymbol{\theta}_{t+1}(t+1)$ as

$$\boldsymbol{\theta}_{t+1}(t+1) = \boldsymbol{\theta}_t(t+1) + b(t+1) \times (\psi^*(X(t+1)) - \psi^*(X^*(t+1))). \qquad (27)$$

Here $b(t+1)$ is another gain function. Usage of the estimator (27) allows avoiding of handling high-dimensional images X, by replacing them with visual features $\psi^*(X)$.

As measurements, used in filtering procedures, it is possible to use low-dimensional representations $r^*(X)$ (21) of vectors $f(X)$ (14) from the RM $\mathbf{M}$, and to construct the estimator $\boldsymbol{\theta}_{t+1}(t+1)$ according to

$$\boldsymbol{\theta}_{t+1}(t+1) = \boldsymbol{\theta}_t(t+1) + d(t+1) \times (\psi^*(X(t+1)) - \psi^*(X^*(t+1))), \qquad (28)$$

where $d(t+1)$ is some gain function.

For choosing the optimal gain functions in (26)-(28), it is necessary to know covariance matrices for deviations between observations and their expected values, as well as between the expected robot position $\boldsymbol{\theta}_t(t+1) = F_t(\boldsymbol{\theta}_t(t))$ (25) and its real pose $\boldsymbol{\theta}(t+1)$. Corresponding covariance matrices can be estimated from the sample $\mathbf{S}_n$ (3), (7) in which robot poses are accurately known.

Acknowledgments. The research was supported solely by the Russian Science Foundation grant (project 14-50-00150).

References

1. Talluri, R., Aggarwal, J.K.: Position estimation techniques for an autonomous mobile robot — A review. In: Handbook of Pattern Recognition and Computer Vision, Chen, C.H., Pau, L.F., Wang, P.S.P.(Eds.), Ch. 4.4, pp. 769–801. Singapore: World Scientific (1993)
2. Borenstein, J.H., Everett, R., Feng, L., Wehe, D.: Mobile robot positioning: Sensors and techniques. J. Robot. Syst. 14, 231–249 (1997)

3. Candy, J.V.: Model-based signal processing. John Wiley & Sons, Inc. (2006)
4. Olson, C.F.: Probabilistic self-localization for mobile robots. IEEE Transactions on Robotics and Automation 16(1), 55–66 (2000)
5. DeSouza, G.N., Kak, A.C.: Vision for mobile robot navigation: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(2), 237267 (2002)
6. Bonin-Font, F., Ortiz, A., Oliver, G.: Visual navigation for mobile robots: A survey. Journal of Intelligent and Robotic Systems 53(3), 263–296 (2008)
7. Krose, B.J.A., Vlassis, N., Bunschoten, R.: Omnidirectional Vision for Appearance-Based Robot Localization. Lecture Notes in Computer Science, v. 2238 "Sensor Based Intelligent Robots", Springer, Berlin Heidelberg, 39–50 (2002)
8. Krose, B.J.A., Vlassis, N., Bunschoten, R., Motomura, Y.: A probabilistic model for appearance-based robot localization. Image and Vision Computing 19, 381–391 (2001)
9. Saito, M., Kitaguchi, K.: Appearance based robot localization using regression models. In: Proceedings of 4th IFAC-Symposium on Mechatronic Systems, v. 2, pp. 584–589 (2006)
10. Hamm, J., Lin, Y., Lee, D.D.: Learning nonlinear appearance manifolds for robot localization. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp. 1239–1244 (2005)
11. Crowley, J.L., Pourraz, F.: Continuity Properties of the Appearance Manifold for Mobile Robot Position Estimation. Image and Vision Computing 19(11), 741-752 (2001)
12. Pauli, J.: Learning-Based Robot Vision. Lecture Notes in Computer Science, vol. 2048 "Principles and Applications", Springer, Heidelberg, 292 pp. (2001)
13. Oore, S., Hinton, G.E., Dudek, G.: A mobile robot that learns its place. Neural Computation 9, 683–699 (1997)
14. Thrun, S.: Bayesian landmark learning for mobile robot localization. Machine Learning 33(1), 41–76 (1998)
15. Krose, B.J.A., Bunschoten, R.: Probabilistic localization by appearance models and active vision. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'99), pp. 2255–2260. Detroit, Michigan (1999)
16. Vlassis, N., Krose, B.J.A.: Robot environment modeling via principal component regression. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS99), pp. 677–682 (1999)
17. Se, S., Lowe, D., Little, J.: Local and global localization for mobile robots using visual landmarks. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2001), pp. 414–420 (2001)
18. Hayet, J., Lerasle, F., Devy, M.: Visual landmarks detection and recognition for mobile robot navigation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), v. 2, pp. 313–318 (2003)
19. Bunschoten, R., Krose, B.J.A.: 3-D scene reconstruction from cylindrical panoramic images. In Proceedings of the 9th International Symposium on Intelligent Robotic Systems (SIRS-2001), pp. 199205 (2001)
20. Gluckman, J., Nayar, S.K.: Ego-motion and omnidirectional cameras. In: Proceedings of the Sixth International Conference on Computer Vision (ICCV 98), pp. 999–1005 (1998)
21. Colin de Verdiere, V., Crowley, J.L.: Local Appearance Space for Recognition of Navigation Landmarks. Journal of Robotics and Autonomous Systems 32(1-2), 61–89 (2000)

22. Dudek, G., Jugessur, D.: Robust Place Recognition using Local Appearance based Methods. In: Proceedings of the International conference on Robotics and Automation (ICRA-00), pp. 1030–1035 (2000)
23. Betke, M., Gurvits, L.: Mobile robot localization using landmarks. IEEE Trans. on Robotics and Automation 13, 251–263 (1997)
24. Sugihara, K.: Some location problems for robot navigation using a single camera. Computer Vision, Graphics, and Image Processing 42, 112–129 (1988)
25. Sim, R., Dudek, G.: Robot positioning using learned landmarks. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 1998), v. 2, pp. 1060–1065 (1998)
26. Friedman, A.: Robot localization using landmarks. The IMA Volumes in Mathematics and its Applications 67(7) "Mathematics in Industrial Problems", Springer, New York, 86–94 (1995)
27. Jogan, M., Leonardis, A.: Robust localization using panoramic view–based recognition. In: Proceedings of the 15th International Conference on Pattern Recognition (ICPR'00), IEEE Computer Society, pp. 136–139 (2000)
28. Vlassis, N., Motomura, Y., Krse, B.J.A.: Supervised Dimension Reduction of Intrinsically Low-Dimensional Data. Neural Computations 14(1), 191–215 (2002)
29. Se, S., Lowe, D., Little, J.: Vision-based global localization and mapping for mobile robots. IEEE Transactions on Robotics 21(3), 364–375 (2005)
30. Cobzas, D., Zhang, H.: Cylindrical panoramic image-based model for robot localization. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1924–1930. Maui, HI (2001)
31. Crowley, J.L., Wallner, F., Schiele, B.: Position estimation using principal components of range data. In: Proceedings of the 1998 IEEE International Conference on Robotics and Automation, vol. 4, pp. 3121–3128 (1998)
32. Jollie, T.: Principal Component Analysis. Springer, New-York (2002)
33. Peres-Neto, P. R., Jackson, D. A., Somers, K.M.: How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics and Data Analysis 49(4), 974–997 (2005)
34. Hardle, W.K., Simar, S.: Canonical Correlation Analysis. In: Applied Multivariate Statistical Analysis, pp. 321–330. Springer–Verlag Berlin Heidelberg (2007)
35. Melzer, T., Reiter, M., Bischof, H.: Appearance models based on kernel canonical correlation analysis. Pattern Recognition, 36(9), 1961–1973 (2003)
36. Skocaj, D., Leonardis, A.: Appearance-based localization using CCA. In: Proceedings of the of the 9th Computer Vision Winter Workshop (CVWW 04), pp. 205–214 (2004)
37. Se, S., Lowe, D., Little, J.: Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. The international Journal of robotics Research 21(8), 735–758 (2002)
38. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
39. Vlassis, N., Motomura, Y., Krose, B.J.A.: Supervised Linear Feature Extraction for Mobile Robot Localization. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 00), vol. 4, pp. 2979–2984 (2000)
40. Saul, L.K., Roweis, S.T.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
41. Wang, K., Wang, W., Zhuang, Y.: Appearance-Based Map Learning for Mobile Robot by Using Generalized Regression Neural Network. Lecture Notes in Computer Science, vol. 4491 "Advances in Neural Networks", Springer-Verlag, Berlin, Heidelberg, 834–842 (2007)

42. Scholkopf, B., Smola, A., Muller, K.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10(5), 1299–1319 (1998)
43. Wu, H., Wu, Y.-X., Liu, C.-A., Yang, G.-T., Qin, S.-Y.: Fast Robot Localization Approach Based on Manifold Regularization with Sparse Area Features. Cognitive Computation 8(5), 856–876 (2016)
44. Do, H.N., Jadaliha, M., Choi, J., Lim, C.Y.: Feature selection for position estimation using an omnidirectional camera. Image and Vision Computing 39, 1-9 (2015)
45. Do, H.N., Choi, J., Lim, C.Y., Maiti, T.: Appearance-based localization using Group LASSO regression with an indoor experiment. In: Proceedings of the 2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM 2015), pp. 984–989 (2015)
46. Do, H.N., Choi, J.: Appearance-based outdoor localization using group lasso regression. In: Proceedings of the ASME Dynamic Systems and Control Conference (DSCC 2015), vol. 3, 8 pp. (2015)
47. Tibshirani, R.: Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society. Series B (Methodological), 73(3), 273–282 (2011)
48. Ribeiro, M.I.: Kalman and extended Kalman filters: Concept, derivation and properties. Institute for Systems and Robotics, Technical Report, 44 pp. (2004)
49. Herbert, B., Andreas, E., Tuytelaars, T., and Gool., L.V.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding 110(3), 346–359 (2008)
50. Obozinski, G., Wainwright, M.J., Jordan, M.I., et al.: Support union recovery in high-dimensional multivariate regression. The Annals of Statistics 39(1), 1–47 (2011)
51. Ma, Y., Fu, Y. (eds.): Manifold Learning Theory and Applications. CRC Press, London (2011)
52. Kuleshov, A.P., Bernstein, A.V.: Manifold Learning in Data Mining Tasks. In: Perner, P. (eds.) MLDM 2014. LNCS, vol. 8556, pp. 119–133. Springer, Heidelberg (2014)
53. Kuleshov, A.P., Bernstein, A.V.: Statistical Learning on Manifold-valued Data. In: Perner, P. (ed.) MLDM 2016. LNCS, vol. 9729, pp. 311–325. Switzerland, Springer International Publishing (2016)
54. Stone, C.J.: Optimal rates of convergence for nonparametric estimators. Annals of statistics 8, 1348–1360 (1980)
55. Stone, C.J.: Optimal global rates of convergence for nonparametric regression. Annals of statistics, v. 10, 1040–1053 (1982)
56. Lee, Jeffrey M.: Manifolds and Differential Geometry. Graduate Studies in Mathematics 107. Providence: American Mathematical Society (2009).
57. Lee, John M.: Introduction to Smooth Manifolds. New York: Springer-Verlag (2003).
58. Bernstein, A.V., Kuleshov, A.P.: Tangent bundle manifold learning via Grassmann & Stiefel eigenmaps. In: arXiv:1212.6031v1 [cs.LG], December 2012, pp. 1–25 (2012)
59. Bernstein, A.V., Kuleshov, A.P.: Manifold Learning: generalizing ability and tangent proximity. Int. J. Softw. Inf. 7(3), 359–390 (2013)
60. Kuleshov, A.P., Bernstein, A.V.: Incremental Construction of Low-dimensional Data Representations. Lecture Notes in Artificial Intelligence, vol. 9896 “Artificial Neural Networks for Pattern Recognition,” pp. 55–67. Springer Heidelberg (2016)

61. Golub, G.H., Van Loan, C.F.: Matrix Computation, 3rd edn. Johns Hopkins University Press, Baltimore, MD (1996)
62. Kuleshov, A.P., Bernstein, A.V.: Regression on High-dimensional Inputs. In: Workshops Proceedings volume of the IEEE International Conference on Data Mining (ICDM 2016), pp. 732–739. USA, IEEE Computer Society (2016)
63. Burnaev, E., Belyaev, M., Kapushev, E.: Computationally efficient algorithm for Gaussian Processes based regression in case of structured samples. Computational Mathematics and Mathematical Physics. 56(4), 499–513 (2016)
64. Burnaev, E., Panov, M., Zaytsev, A.: Regression on the Basis of Nonstationary Gaussian Processes with Bayesian Regularization. Journal of Communications Technology and Electronics. 61(6), 661–671 (2016)
65. Burnaev, E., Zaytsev, A. Surrogate modeling of mutlifidelity data for large samples. Journal of Communications Technology and Electronics. 60(12), 1348–1355 (2016)